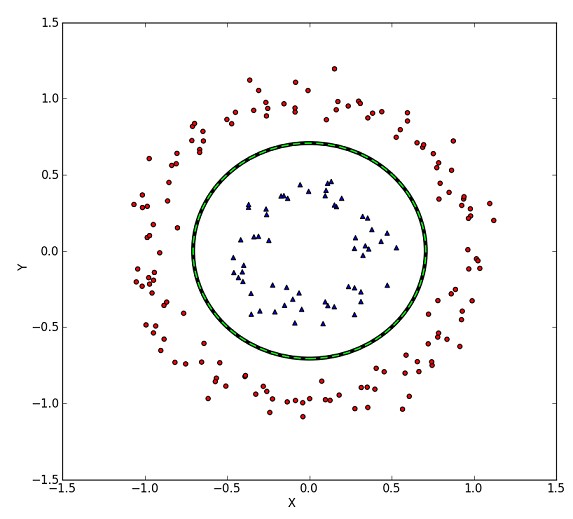
第22课：SVM——非线性 SVM 和核函数

### 非线性分类问题

遇到分类问题的时候，最理想的状态，当然是样本在向量空间中都是线性可分的，我们可以清晰无误地把它们分隔成不同的类别——****线性可分 SVM****。

如果实在不行，我们可以容忍少数不能被正确划分，只要大多数线性可分就好——****线性 SVM****。

可是，如果我们面对的分类问题，根本就是****非线性****的呢？比如像下面这样：



图中红色的点是正类样本，蓝色的点是负类样本。通过我们的观察可知，它们之间的界限是很分明的，用图中绿色的圈本来可以把它们完全分开。

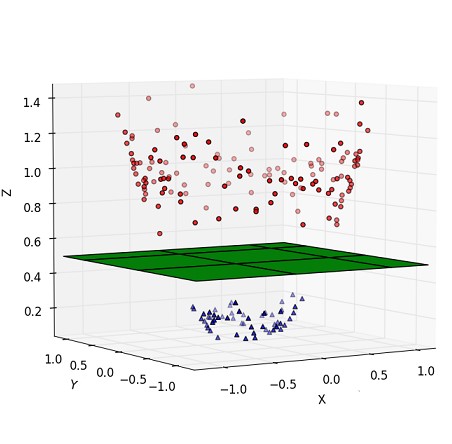
很可惜，“圆圈”在二维空间里无法用线性函数表示，也就是说这些样本在二维空间里根本线性不可分。所以，无论是线性可分 SVM 还是线性 SVM，都无法在这些样本上良好工作。

这可怎么办呢？难道，这种情况我们就处理不了了？

并不是！

我们可以想个办法，让这些在二维空间中线性不可分的样本，****在更高维度的空间里线性可分****。

比如说，如果我们能把上图中那些正负类的样本映射到三维空间中，并且依据不同的类别给它们赋予不一样的高度值——zz 轴取值（就像下图这样），那么不就线性可分了嘛。



如此一来，在二维空间团团转的正负例，在三维空间中分为两层，中间用一个超平面，就可以完美分隔了。

### 非线性 SVM

#### 非线性 SVM 分隔超平面

对于在有限维度向量空间中线性不可分的样本，我们将其映射到更高维度的向量空间里，再通过间隔最大化的方式，学习得到支持向量机，就是****非线性 SVM****。

我们将样本映射到的这个更高维度的新空间叫做****特征空间****。

注意：如果是理想状态，样本从原始空间映射到特征空间后直接就成为线性可分的，那么接下来的学习是可以通过硬间隔最大化的方式来学的。

不过，一般的情况总没有那么理想，因此，通常情况下，我们还是按照软间隔最大化，在特征空间中学习 SVM。

简单理解就是：****非线性 SVM = 核技巧 + 线性 SVM****。

我们用向量 xx 表示位于原始空间中的样本，ϕ(x)ϕ(x) 表示 xx 映射到特征空间之后的新向量。

则****非线性 SVM****对应的****分隔超平面为：****f(x)=wϕ(x)+bf(x)=wϕ(x)+b。

#### 非线性 SVM 的对偶问题

套用上一篇线性 SVM 的对偶问题，此处****非线性 SVM 的对偶问题****就变成了：

min(12∑mi=1∑mj=1αiαjyiyjϕ(xi)⋅ϕ(xj)−∑mi=1αi)min(12∑i=1m∑j=1mαiαjyiyjϕ(xi)⋅ϕ(xj)−∑i=1mαi)

s.t.∑mi=1αiyi=0s.t.∑i=1mαiyi=0

0⩽αi⩽C,i=1,2,...,m0⩽αi⩽C,i=1,2,...,m

大家可以看到，****和线性 SVM 唯一的不同****就是：之前的 xixi****与****xjxj****的内积（点乘） 变成了****ϕ(xi)ϕ(xi)****与****ϕ(xj)ϕ(xj)****的内积****。

#### 核函数

对于有限维的原始空间，一定存在更高维度的空间，使得前者中的样本映射到新空间后可分。但是新空间（特征空间）的维度也许很大，甚至可能是无限维的。这样的话，直接计算 ϕ(xi)⋅ϕ(xj)ϕ(xi)·ϕ(xj) 就会很困难。

为了避免计算 ϕ(xi)ϕ(xi) 和 ϕ(xj)ϕ(xj) 的内积，我们需要设置一个新的函数——k(xi,xj)k(xi,xj)：

k(xi,xj)=ϕ(xi)⋅ϕ(xj)k(xi,xj)=ϕ(xi)⋅ϕ(xj)

原始空间中的两个样本 xixi 和 xjxj 经过 k(⋅,⋅)k(·,·) 函数计算所得出的结果，是它们在特征空间中映射成的新向量的内积。

如此一来，我们就不必真的计算出 ϕ(xi)ϕ(xi) 点乘 ϕ(xj)ϕ(xj) 的结果，而可以直接用 k(⋅,⋅)k(·,·) 函数代替它们。

我们把这个 k(⋅,⋅)k(·,·) 函数叫做核函数。现在我们给出它的正式定义：

设 XX 为原始空间（又称输入空间），HH 为特征空间（特征空间是一个带有内积的完备向量空间，又称完备内积空间或希尔伯特空间）。

如果存在一个映射： X×XX×X ，使得对所有 xi,xj∈Xxi,xj∈X，函数 k(xi,xj)k(xi,xj) 满足条件：k(xi,xj)=ϕ(xi)⋅ϕ(xj)k(xi,xj)=ϕ(xi)⋅ϕ(xj)，即 k(⋅,⋅)k(·,·) 函数为输入空间中任意两个向量映射到特征空间后的内积。

则称 k(⋅,⋅)k(·,·) 为核函数，ϕ(⋅)ϕ(·) 为映射函数。

#### 运用核技巧求解非线性 SVM 的对偶问题

有了核函数，我们就可以将非线性 SVM 的对偶问题写成：

min(12∑mi=1∑mj=1αiαjyiyjk(xi,xj)−∑mi=1αi)min(12∑i=1m∑j=1mαiαjyiyjk(xi,xj)−∑i=1mαi)

s.t.∑mi=1αiyi=0s.t.∑i=1mαiyi=0

0⩽αi⩽C,i=1,2,...,m0⩽αi⩽C,i=1,2,...,m

之后的求解过程与线性 SVM 一致：先根据对偶问题求解 αα，再根据 αα 的结果计算 ww，然后根据支持向量求解 bb，在此就不赘述了。

相应的，最终求出的非线性 SVM 在特征空间的最大分隔超平面也就成了：

f(x)=wϕ(x)+b=∑mi=1αiyiϕ(xi)⋅ϕ(x)+b=∑mi=1αiyik(x,xi)+bf(x)=wϕ(x)+b=∑i=1mαiyiϕ(xi)⋅ϕ(x)+b=∑i=1mαiyik(x,xi)+b

上述运用核函数求解的过程，称为核技巧。

### 核函数的性质

如果我们已经知道了映射函数是什么，当然可以通过两个向量映射后的内积直接求得核函数。

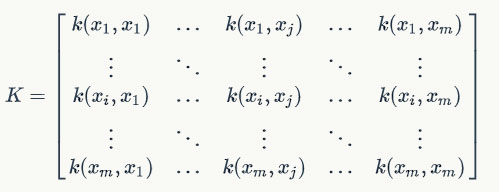
但是，在我们还不知道映射函数本身是什么的时候，有没有可能直接判断一个函数是不是核函数呢？

换句话说，****一个函数需要具备怎样的性质，才是一个核函数****？

以下是****函数****k(⋅,⋅)k(·,·)****可以作为核函数的充分必要条件****：

设 XX 是输入空间，k(⋅,⋅)k(·,·) 是定义在 X×XX×X 上的对称函数，当且仅当任意 x1,x2,...,xm∈Xx1,x2,...,xm∈X， 核矩阵 KK（如下所示）总是半正定时，k(⋅,⋅)k(·,·) 就可以作为核函数使用。

核矩阵：



注意：这里有几个线性代数的基础概念，请见下。

对称函数：函数的输出值不随输入变量的顺序改变而改变的函数叫做对称函数。

因为输入空间中的 xi和xjxi和xj 是向量，特征空间中的 ϕ(xi)和ϕ(xj)ϕ(xi)和ϕ(xj) 也是向量，k(xi,xj)k(xi,xj) 表示特征空间向量的内积，而两个向量的内积并不因为其顺序变化而变化。因此，k(xi,xj)=k（xj,xi)k(xi,xj)=k（xj,xi)，即 k(⋅,⋅)(·,·) 为对称函数。

相应的，核矩阵 K 为对称矩阵。

半正定矩阵：一个 n×nn×n 的实对称矩阵 MM 为半正定，当且仅当对于所有非零实系数向量 zz，均有： zTMz⩾0zTMz⩾0。

其中“非零实系数向量”的含义是：一个向量中所有元素均为实数，且其中至少有一个元素值非零。

### 核函数的种类

要知道，非线性 SVM 的****关键****在于将输入空间中线性不可分的样本映射到线性可分的特征空间中去。特征空间的好坏直接影响到了 SVM 的效果。

****如何选择核函数****也就成了一个关键性的问题。

虽然我们已经学习了核函数的定义和性质，但让我们凭空去自己构建一个核函数出来，还是一个非常困难的任务。

即使真得构造出来了，这个核函数在具体问题上的效果如何，还需要通过大量测试来证明，很可能费了好大劲，最后效果还不理想。

好在前人已经给我们留下来一些经历了长久磨练的常用核函数，让我们可以直接拿来用。接下来，我们分别看下这些核函数。

#### 线性核（Linear Kernel）

k(xi,xj)=xTixjk(xi,xj)=xiTxj

使用时无须指定参数（Parameter），它直接计算两个输入向量的内积。经过线性核函数转换的样本，特征空间与输入空间重合，相当于并没有将样本映射到更高维度的空间里去。

很显然这是最简单的核函数，实际训练、使用 SVM 的时候，在不知道用什么核的情况下，可以先试试线性核的效果。

#### 多项式核（Polynomial Kernel）

k(xi,xj)=(γxTixj+r)d,γ>0,d⩾1k(xi,xj)=(γxiTxj+r)d,γ>0,d⩾1

需要指定三个参数：γγ，rr 和 dd。

这是一个不平稳的核，适用于数据做了归一化（Normalization，参见本文最后一节）的情况。

#### RBF 核（Radial Basis Function Kernel）

k(xi,xj)=exp(−γ||xi−xj||2),γ>0k(xi,xj)=exp⁡(−γ||xi−xj||2),γ>0

RBF 核又名高斯核 （Gaussian Kernel），是一个核函数家族。它会将输入空间的样本以非线性的方式映射到更高维度的空间（特征空间）里去，因此它可以处理类标签和样本属性之间是非线性关系的状况。

它有一个参数：γγ，这个参数的设置非常关键！

如果设置过大，则整个 RBF 核会向线性核方向退化，向更高维度非线性投影的能力就会减弱；但如果设置过小，则会使得样本中噪音的影响加大，从而干扰最终 SVM 的有效性。

不过相对于多项式核的3个参数，RBF 核只有一个参数需要调，还是相对简单的。

当线性核效果不是很好时，可以用 RBF 试试。或者，很多情况下可以直接使用 RBF。

著名的 LIBSVM（台湾大学林智仁/Lin Chih-Jen 教授等设计开发的一款简单、易用、快速有效的 SVM/SVR 支持包）的默认核函数，就是 RBF 核。

#### Sigmoid 核（Sigmoid Kernel）

k(xi,xj)=tanh(γxTixj+r)k(xi,xj)=tanh(γxiTxj+r)

有两个参数，γγ 和 rr，在某些参数设置之下，Sigmoid 核矩阵可能不是半正定的，此时 Sigmoid 核也就不是有效的核函数了。因此参数设置要非常小心。

整体而言，Sigmoid 核并不比线性核或者 RBF 核更好。但是，当参数设置适宜时，它会有不俗的表现。

在具体应用核函数时，最好针对具体问题参照前人的经验。

### 构建自己的核函数

除了常见的核函数，我们还可以根据以下规律进行核函数的组合。

与正数相乘： k(⋅,⋅)k(·,·) 是核函数，对于任意正数（标量）α⩾0,αk(⋅,⋅)α⩾0,αk(·,·) 也是核函数。

与正数相加： k(⋅,⋅)k(·,·) 是核函数，对于任意正数（标量）α⩾0,α+k(⋅,⋅)α⩾0,α+k(·,·) 也是核函数。

线性组合： k(⋅,⋅)k(·,·) 是核函数，则如下线性组合也是核函数：∑ni=1αiki(⋅,⋅),αi⩾0∑i=1nαiki(⋅,⋅),αi⩾0。

乘积： k1(⋅,⋅)k1(·,·) 和 k2(⋅,⋅)k2(·,·) 都是核函数，则它们的乘积——k1(⋅,⋅)k2(⋅,⋅)k1(·,·)k2(·,·)——也是核函数。

正系数多项式函数： 设 PP 为实数域内的正系数多项式函数，k(⋅,⋅)k(·,·) 是核函数，则 P(k(⋅,⋅))P(k(·,·)) 也是核函数。

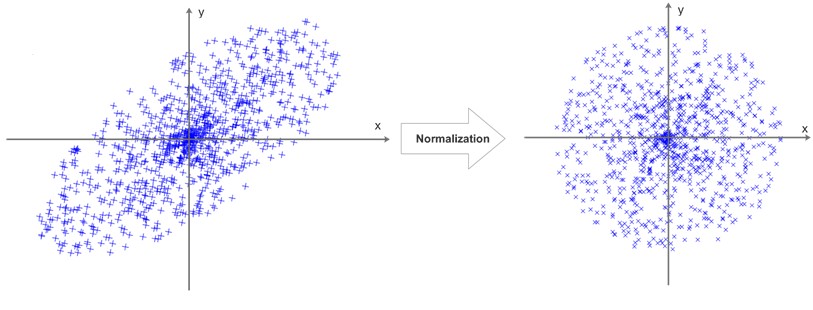
指数函数： k(⋅,⋅)k(·,·) 是核函数，则 exp(k(⋅,⋅))exp⁡(k(·,·)) 也是核函数。

掌握了这些规律，我们也可以尝试根据需要构建自己的核函数。

### 数据归一化（Data Normalization）

数据归一化是一种数据处理方法，具体所做的就是对取值范畴不同的数据进行归一化处理，使它们处在同一数量级上。

最常见的，就是把各种数据都变成 (０,１)(０,１) 之间的小数。



上图是一个归一化过程在二维中的直观显示。

大家可以看到， “扁长”的数据分布，经过归一化处理之后，变成了一个“正圆”。

常用的归一化算法有：

（1）线性转换

x′=(x−min)(max−min)x′=(x−min)(max−min)

（2）标准分

x′=(x–μ)γx′=(x–μ)γ

原本的 xx 符合正态分布，μμ 为其分布的均值，而 γγ 为分布的方差。